The computation of C-C and N-N bond dissociation energies for singly, doubly, and triply bonded systems

Stephen R. Langhoff and Charles W. Bauschlicher, Jr.
NASA Ames Research Center
Moffett Field, CA 94035

and

Peter R. Taylor
ELORET Institute†
Palo Alto, CA 94303

Abstract

The bond dissociation energies \( D_e \) of \( \text{C}_2\text{H}_2, \text{C}_2\text{H}_4, \text{C}_2\text{H}_8, \text{N}_2, \text{N}_2\text{H}_2, \) and \( \text{N}_2\text{H}_4 \) are studied at various levels of correlation treatment. The convergence of \( D_e \) with respect to the one-particle basis is studied at the single-reference modified coupled-pair functional (MCPF) level. At all levels of correlation treatment, the errors in the bond dissociation energies increase with the degree of multiple bond character. The multireference configuration-interaction (MRCI) \( D_e \) values, corrected for an estimate of higher excitations, are in excellent agreement with those determined using the size-extensive averaged coupled-pair functional (ACPF) method. We find that the full-valence complete-active-space self-consistent-field (CASSCF)/MRCI calculations are reproduced very well by MRCI calculations based on a CASSCF calculation that includes in the active space only those electrons involved in the C-C or N-N bonds. To achieve chemical accuracy (1 kcal/mole) for the \( D_e \) values of the doubly bonded species \( \text{C}_2\text{H}_4 \) and \( \text{N}_2\text{H}_2 \) requires one particle basis sets including up through \( h \) angular momentum functions \((l=5)\) and a multireference treatment of electron correlation: still higher levels of calculation are required to achieve chemical accuracy for the triply bonded species \( \text{C}_2\text{H}_2 \) and \( \text{N}_2 \).

† Mailing address: NASA Ames Research Center, Moffett Field, CA 94035
I. Introduction

The modelling of combustion processes requires accurate bond energies for many molecular species. Recent work [1-3] has shown that C–C and C–H single bond dissociation energies can be straightforwardly computed using current \textit{ab initio} methods. However, for electron-dense systems such as C$_2$H$_2$ and N$_2$, it is well known that extensive one-particle basis sets and correlation treatments are required to approach chemical accuracy for the binding energy. The greater degree of difficulty in treating multiply-bonded systems can potentially complicate the determination of heats of formation. For example the Haber process for the synthesis of ammonia:

\[ \text{N}_2 + 3\text{H}_2 \rightarrow 2\text{NH}_3 \]  \hspace{1cm} (1)

involves breaking a triple bond and three H–H single bonds and forming six polar N–H single bonds. Thus if the triple bond is treated less accurately than the single bonds, this can potentially introduce an error into the calculated heat of formation. Yet calculations using quadratic configuration-interaction methods in conjunction with extrapolation to the complete basis set limit (CBS-QCI) [4] give excellent agreement with experiment [5]. On the other hand, the computation of even C–H single bond dissociation energies is complicated if in the process of breaking the bond the degree of multiple bond character elsewhere in the molecule changes significantly [3]. Thus it is important to understand the limitations of various theoretical methods as to their ability to treat equivalently single, double, and triple bonds.

In this work we determine the binding energies of the singly bonded species C$_2$H$_6$ and N$_2$H$_4$, the doubly bonded species C$_2$H$_4$ and N$_2$H$_2$, and the triply bonded species C$_2$H$_2$ and N$_2$ at various levels of theory. One goal of our study is to calibrate the single-reference-based MCPF approach, as this can be applied to relatively large systems using high quality one-particle basis sets. For our multireference calculations, we consider whether it is necessary or desirable to perform full valence CASSCF calculations, or whether use of a smaller active space including only those electrons affected in the bond-breaking process is adequate. In the next section we discuss the details of the theoretical approaches employed in this work. In section III we compare the calculated binding energies at various levels of theory with experiment, as accurate $D_0$ values and zero-point corrections are available for these species. Our conclusions are presented in section IV.
II. Methods

Except for basis set calibration studies, we have used atomic natural orbital (ANO) basis sets [6] of the form (13s 8p 6d 4f)/[4s 3p 2d 1f] for carbon and nitrogen and (8s 6p 4d)/[4s 2p 1d] for hydrogen. These basis sets are described in more detail in Refs. 6 and 7. For all systems we carried out MCPF calculations after adding an additional s, p, d and f ANO function, as well as a (2g)/[1g] ANO function to the heavy atom basis. In addition, for the doubly and triply bonded species we added an additional s – g ANO and a primitive h function. The orbital exponents of the h function added to C and N are 1.24 and 1.728, respectively. Thus the largest contracted basis sets employed for C and N were [6s 5p 4d 3f 2g lh]. We also expanded the hydrogen basis to a nearly complete [6s 5p 2d lh] set [8] for calculations on C₂H₄. Extensive basis set studies were also considered for N₂ in earlier work [9,10]. The smaller [4s 3p 2d 1f] basis sets, however, are entirely adequate for addressing the limitations of the various correlation approaches. Only the pure spherical harmonic components of the basis functions were used in the calculations.

Both single- and multireference-based correlation approaches were used to determine the bond dissociation energies. For our single-reference-based approach, we use the approximately size-extensive MCPF method [11], which has been shown to provide accurate dissociation energies for single bonds provided both the molecule and its fragments are well described by an SCF reference. This is the case for all of the systems considered in this work.

Our multireference-based methods include the complete-active-space self-consistent-field (CASSCF) multireference configuration-interaction (MRCI) and averaged coupled-pair functional (ACPF) methods [12]. In the initial choice of the active space we included all of the bonding and antibonding orbitals and electrons. Thus all electrons were correlated except the nominally 1s electrons on C and 1s and 2s electrons on N. However, for C₂H₆ and N₂H₄ this resulted in a very large CASSCF-CI expansion. For example, for C₂H₄ the full-valence calculation corresponds to 14 electrons in 14 orbitals. As this was very time consuming we carried out a 10 in 10 calculation instead, where one orbital in each of the a₁g and a₂u symmetries of the D₃d point group were moved to the inactive space and the corresponding antibonding orbitals were moved to the secondary space. These orbitals correspond to the symmetric and antisymmetric (with respect to the C-C bond)
combinations of the CH bond orbitals that are totally symmetric on each fragment. We believe that MRCI calculations based on either a 14 in 14 or 10 in 10 CASSCF treatment will give very similar results. An analogous reduction of the \( \text{N}_2\text{H}_4 \) calculation from 10 in 10 to 6 in 6 was also performed. The second choice of the CASSCF active space included only those electrons involved in either the C-C or N-N bonds. This was possible as these orbitals could be clearly distinguished from those involved in the bonds to hydrogen. For the MRCI and ACPF calculations it was not possible to include all single and double excitations from the full-valence CASSCF wave function, as this led to prohibitively long expansions. Thus the reference lists were restricted to include all occupations for which the absolute value of the coefficient of any one of its component spin couplings exceeded a designated threshold (generally 0.05) in the CASSCF wave function. An advantage of the smaller CASSCF active spaces is that no reference selection was required. That is, all MRCI and ACPF calculations involving the smaller active spaces include the full CASSCF as the reference wave function. All electrons, except the 1s electrons on carbon and nitrogen, were correlated in the MCPF, MRCI and ACPF calculations. The effect of higher excitations in the MRCI calculations were estimated using the multireference analog of the Davidson correction, denoted \(+Q\). The reference wave function for the ACPF calculations was expanded until it included all configurations with coefficients greater than 0.05 in the final ACPF wave function.

The \( D_e \) values were computed as the energy difference between the equilibrium geometry and a dissociated supermolecule structure. Experimental geometries were used for both the molecules and fragments [13-15]. For the MCPF treatment of \( \text{C}_2\text{H}_2 \), it was necessary to dissociate to two CH fragments in the excited \( 4\Sigma^- \) state, and then to correct to ground state fragments using the known \( 4\Sigma^- - 2\Pi \) separation for \( \text{CH} \) [13]. For \( \text{N}_2\text{H}_2 \) (diimide), we used the experimental geometry for the trans structure. For \( \text{N}_2\text{H}_4 \) we used the gauche equilibrium structure [14,15]. All calculations were carried out using the MOLECULE-SWEDEN [16] program system on the NASA Ames Research Center Central Computing Facility Y-MP/832.

III. Results and Discussion

The \( D_e \) values as a function of correlation treatment are summarized in Table I for the six molecular systems considered in this work. An important conclusion that can be drawn from the results in Table I is the excellent agreement of the
MRCI+Q and ACPF results based on both the full-valence and smaller CASSCF active spaces. This is particularly true of the C₂H₂ₙ species where the agreement is nearly exact. In fact, there is more variation of the full-valence MRCI results with selection threshold than between the MRCI calculations based on the two choices of active space. The excellent agreement between the MRCI+Q and ACPF \( D_e \) values indicates that the contribution from higher excitations has been accounted for satisfactorily. Probably our best results are the ACPF values using the smaller active space, as the ACPF incorporates the effects of higher excitations in a more rigorous manner and the smaller active space avoids errors inherent in reference selection. In comparing these ACPF values with the MCPF values, we see that the discrepancy increases with the degree of multiple bond character. For single bonds the error is between 0-1 kcal/mole, for double bonds it is of the order 3-4 kcal/mole, and for triple bonds it is of the order of 6 kcal/mole. This is consistent with expectations that higher excitations are more important in describing multiply bonded systems.

Our theoretical results in Table I are of course limited in accuracy by approximations in both the one- and \( n \)-particle expansions. Previous studies [10,17] for \( \text{N}_2 \) indicate that the ACPF method gives binding energies within 1-2 kcal/mole of the full configuration-interaction (FCI) limit. We expect that the ACPF treatment will be even closer to the FCI limit for the singly and doubly bonded species, since higher excitations will be less important than for \( \text{N}_2 \). However, the use of a [4s 3p 2d 1f] ANO basis set for C and N results in a substantially larger underestimation of the \( D_e \) value than does the use of the ACPF method for accounting for electron correlation. To assess these limitations, we have carried out basis set calibration studies at the MCPF level—see Table II. Two trends are apparent from the results in Table II. First, the magnitude of the basis set corrections increase with the degree of multiple bondedness, and second, the corrections are significantly larger for the nitrogen species. It is interesting to note that the basis set requirements are also larger for N-H as compared with C-H bonds. For example, at the MCPF level, increasing the basis set from [4s 3p 2d 1f] to [5s 4p 3d 2f 1g] increases the first N-H bond dissociation energy of \( \text{NH}_3 \) by 1.2 kcal/mole, whereas it increases the C-H bond dissociation energy of \( \text{CH}_4 \) by only 0.5 kcal/mole.

As can be seen from Table II, the effect of expanding the basis from [5s 4p 3d 2f 1g] to [6s 5p 4d 3f 2g 1h] is on the average about 0.35 that of expanding it from [4s 3p 2d 1f] to [5s 4p 3d 2f 1g]. Also, the contribution from adding a single
The function is about 0.30-0.35 that of adding the first ANO g function. We have made an estimate for the remaining basis set incompleteness by assuming that further basis set saturation and the addition of still higher angular momentum basis functions increases $D_e$ in a similar manner. This is essentially a geometric approximation: the correlation energy itself converges more slowly than this, but the convergence of the binding energy is uncertain. Our approximation will certainly be an underestimate of the remaining basis set correction, especially since it does not account for further saturation in the primitive basis: note that expanding the N basis from $(13s\ 8p\ 6d\ 4f\ 2g\ 1h)/[6s\ 5p\ 4d\ 3f\ 2g\ 1h]$ to $(18s\ 13p\ 6d\ 5f\ 4g\ 3h)/[6s\ 5p\ 4d\ 3f\ 2g\ 1h]$ increases the $D_e$ of N$_2$ by 0.2 kcal/mole at the MRCI level [10]. The effect of basis set saturation in N$_2$ is very similar at the MCPF and MRCI levels of correlation treatment, so that the MCPF approach is expected to be adequate for evaluating basis sets effects on the $D_e$ of the other molecules considered in this work. Our extrapolation procedure does not account for improvements in the hydrogen basis. However, expanding the hydrogen basis set from $[4s\ 2p\ 1d]$ to $[6s\ 5p\ 2d\ 1f]$, in conjunction with a $[5s\ 4p\ 3d\ 2f\ 1g]$ ANO basis set for carbon, increased the $D_e$ of C$_2$H$_4$ by only 0.1 kcal/mole. This very small increase in $D_e$ is not surprising, of course, since the hydrogen basis functions contribute very little to the C-C bond breaking process. The values in Table II in the line denoted “total” are our estimate of expanding the heavy atom basis set from $[4s\ 3p\ 2d\ 1f]$ to the basis set limit at the MCPF level of correlation treatment. We use these values in Table III to correct our ACPF $D_e$ values in the $[4s\ 3p\ 2d\ 1f]$ basis before comparing with experiment and the values deduced using the G1 approach [18]. We have not included basis set superposition error (BSSE) corrections, as these are expected to be small. For example, the BSSE determined using the counterpoise correction with the full ghost basis was 0.28 kcal/mole for N$_2$ in the $[6s\ 5p\ 4d\ 3f\ 2g\ 1h]$ basis [10]. This correction is expected to be less for the carbon species and for the singly and doubly bonded nitrogen species.

To facilitate comparison with experiment, we have converted our $D_e$ values to $D_0$ values using the experimental vibrational frequencies [19,20]. The $D_0$ values denoted theory in Table III are our computed $D_e$ values at the ACPF level (using the smaller active space) in the $[4s\ 3p\ 2d\ 1f]/[4s\ 2p\ 1d]$ ANO basis corrected for zero-point and by our estimate for extrapolation to the basis set limit. Note that we have deduced the experimental binding energy of N$_2$H$_2$ using a heat of formation, $\Delta H_f^{298}$.
for trans-diimide of 36±2 kcal/mole [21]. Our calculated $D_0$ value is inconsistent with the two larger experimental values [22,23] for $\Delta H^0_f$, as well as the calculated value of Casewit and Goddard [24].

Our theoretical $D_0$ values in Table III tend to be smaller than either experiment [13,19,21,25] or the values determined using the G1 approach [18]. Overall our $D_0$ values for the carbon species are in better agreement with experiment than are the G1 values, whereas the opposite is true for the nitrogen species. Our values are almost certainly lower bounds, as the ACPF method tends to underestimate slightly (at least for the multiply bonded species) the correlation contribution to $D_e$, and our basis set extrapolation procedure does not account for saturation of the primitive basis.

Considering that our theoretical values in Table III are probable lower bounds, the correct $D_0$ value for $C_2H_4$ probably lies in the upper half of the experimental error bars. Most of the remaining 2 kcal/mole discrepancy with experiment for the C-C bond dissociation energy of $C_2H_2$ can probably be attributed to limitations in the ACPF method. Our theoretical $D_0$ value of $N_2H_4$ is smaller than the experimental value, but within the error bars that are relatively large due to the uncertainty in the heat of formation of NH$_2$. The G1 value is also smaller than experiment, but it should be noted that the G1 value is determined using a theoretically determined zero-point correction of −9.5 kcal/mole, instead of the experimental value of −7.2 kcal/mole. Thus the G1 $D_0$ value would be larger than experiment, in analogy with $C_2$H$_6$, if the experimental zero-point correction [20] was used instead.

Petersson and Mantzaris [4] have computed the heat of formation of ammonia using the CBS-QCI model. Their value of −8.83±0.5 kcal/mole is in good agreement with the accurate experimental value [5] of −9.34±0.01 kcal/mole. They present a careful analysis of why the value of −6.52 kcal/mole obtained with the generally very reliable G1 model [26] is in error by nearly 3 kcal/mole. Since the Haber process for the synthesis of ammonia involves breaking a triple bond in N$_2$ and three single H-H bonds and subsequently forming polar single bonds in NH$_3$, we considered how reliably the MCPF approach could determine the heat of formation using the [4s 3p 2d 1f/4s 2p 1d] ANO basis. Using Petersson and Mantzaris's value of 10.31 kcal/mole for the zero-point correction, we compute a $\Delta H^0_f$ value of −6.97 and −8.25 kcal/mole at the SCF and MCPF levels. For comparison in
this basis set the CCSD and CCSD(T) methods produce $\Delta H_f^0$ values of $-8.83$ and $-7.81$ kcal/mole, respectively. Thus the MCPF value lies between the CCSD and CCSD(T) values. Triple excitations tend to decrease $\Delta H_f^0$, because they are more important for $N_2$. Expanding the basis set to $[5s\ 4p\ 3d\ 2f\ 1g/4s\ 2p\ 1d]$, decreases the MCPF value to $-9.24$ kcal/mole, in near perfect agreement with experiment. Thus there is a relatively small effect of basis set saturation, which in large part arises from a cancellation of errors that occurs when the process is formulated as an isogyric reaction. As there is very little error in the bond energy of $H_2$, obtaining the correct heat of reaction for (1) (and thus the correct heat of formation for $NH_3$) relies on a cancellation of the error in the triple bond of $N_2$ and the error in six N-H bonds. This cancellation is nearly exact at the MCPF level in the $[5s\ 4p\ 3d\ 2f\ 1g]$ basis, and based on our basis set studies we expect that the $\Delta H_f^0$ value of $NH_3$ will also be accurate at the MCPF level in the basis set limit. This is consistent with the rather good result obtained by Petersson and Mantzaris [4] with the CBS-QCI approach. We should also note that it should be possible to determine $\Delta H_f^0$ for $CH_4$ accurately by a similar procedure. The $D_e$ of $C_2$ in the $[4s\ 3p\ 2d\ 1f]$ basis at the MCPF level is about 11 kcal/mole less than experiment [27], and the C-H bond strength at this level is 1.5 kcal/mole less than experiment [25]. Thus the errors are comparable on both sides of the reaction $4H_2 + C_2 \rightarrow 2CH_4$.

IV. Conclusions

We have carried out calculations for the singly bonded species, $C_2H_6$ and $N_2H_4$, the doubly bonded species, $C_2H_4$ and $N_2H_2$, and the triply bonded species $C_2H_2$ and $N_2$ at various levels of correlation treatment to assess how the errors in the binding energies vary with the degree of multiple bondedness. We find that the errors in the single-reference approach as compared with the multireference ACPF approach increase from 0-1 kcal/mole for single bonds to 3-4 kcal/mole for double bonds to approximately 6 kcal/mole for triple bonds. In addition, the one-particle basis set requirements also increase substantially with the degree of multiple bondedness. Thus while it may be possible to compute binding energies to chemical accuracy by direct ab initio methods for singly bonded species, it will be necessary to either employ very large one-particle basis sets (including at least $h$ functions) or an extrapolation procedure to account for the remaining basis set incompleteness, such as those performed in the G1 (and G2) approaches or the CBS-QCI model,
in order to obtain quantitative results for the $D_e$ values of multiple bonds. The multireference calculations performed in this work underscore the desirability of restricting the CASSCF space to include only those orbitals and electrons involved in the bond breaking process.
REFERENCES


Table I. $D_e$ values (kcal/mole) as a function of the level of correlation treatment.

<table>
<thead>
<tr>
<th>System</th>
<th>Active space</th>
<th>$D_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CASSCF</td>
<td>MRCI</td>
</tr>
<tr>
<td>$\text{C}_2\text{H}_2$</td>
<td>$10\times10$</td>
<td>232.0</td>
</tr>
<tr>
<td></td>
<td>$6\times6$</td>
<td>236.4</td>
</tr>
<tr>
<td>$\text{C}_2\text{H}_4$</td>
<td>$12\times12$</td>
<td>167.2</td>
</tr>
<tr>
<td></td>
<td>$4\times4$</td>
<td>176.3</td>
</tr>
<tr>
<td>$\text{C}_2\text{H}_6$</td>
<td>$10\times10^a$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$2\times2$</td>
<td>80.8</td>
</tr>
<tr>
<td>$\text{N}_2$</td>
<td>$6\times6$</td>
<td>203.2</td>
</tr>
<tr>
<td>$\text{N}_2\text{H}_2$</td>
<td>$8\times8$</td>
<td>114.3</td>
</tr>
<tr>
<td></td>
<td>$4\times4$</td>
<td>122.0</td>
</tr>
<tr>
<td>$\text{N}_2\text{H}_4$</td>
<td>$6\times6^a$</td>
<td>64.8</td>
</tr>
<tr>
<td></td>
<td>$2\times2$</td>
<td>51.8</td>
</tr>
</tbody>
</table>

$^a$ See the text.

$^b$ Based on SCF orbitals.
Table II. MCPF $D_e$ values (kcal/mole) as a function of the heavy atom basis set.

<table>
<thead>
<tr>
<th>Basis</th>
<th>Molecule</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_2H_2$</td>
</tr>
<tr>
<td>[4s 3p 2d 1f]</td>
<td>223.03</td>
</tr>
<tr>
<td>[5s 4p 3d 2f]</td>
<td>225.35</td>
</tr>
<tr>
<td>[5s 4p 3d 2f 1g]</td>
<td>227.52</td>
</tr>
<tr>
<td>[6s 5p 4d 3f 2g]</td>
<td>228.26</td>
</tr>
<tr>
<td>[6s 5p 4d 3f 2g 1h]</td>
<td>228.88</td>
</tr>
<tr>
<td>$\infty^d$</td>
<td>229.47</td>
</tr>
<tr>
<td>Total</td>
<td>6.44</td>
</tr>
</tbody>
</table>

\( ^a\) The hydrogen basis is [4s 2p 1d].
\( ^b\) The total basis set correction is based on the difference between the $D_e$ values in the [4s 3p 2d 1f] and [5s 4p 3d 2f 1g] basis sets and the total correction for the corresponding doubly bonded system.
\( ^c\) The $D_e$ value increases to 176.85 when the hydrogen basis set is expanded to [6s 5p 2d 1f].
\( ^d\) Estimate based on a geometric extrapolation.
Table III. Comparison of the calculated C-C and N-N bond strengths (kcal/mole) with experiment.

<table>
<thead>
<tr>
<th>System</th>
<th>$D_e^2$</th>
<th>Corrections</th>
<th>$D_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Zero-point</td>
<td>Basis set$^b$</td>
</tr>
<tr>
<td>$C_2H_2$</td>
<td>228.3</td>
<td>-8.1</td>
<td>+6.4</td>
</tr>
<tr>
<td>$C_2H_4$</td>
<td>177.3</td>
<td>-9.7</td>
<td>+3.6</td>
</tr>
<tr>
<td>$C_2H_6$</td>
<td>94.4</td>
<td>-9.0</td>
<td>+1.6</td>
</tr>
<tr>
<td>$N_2$</td>
<td>215.7</td>
<td>-3.4</td>
<td>+9.9</td>
</tr>
<tr>
<td>$N_2H_2$</td>
<td>122.8</td>
<td>-7.8</td>
<td>+5.7</td>
</tr>
<tr>
<td>$N_2H_4$</td>
<td>67.5</td>
<td>-7.2</td>
<td>+3.5</td>
</tr>
</tbody>
</table>

$^a$ ACPF results obtained using the [4s 3p 2d 1f]/[4s 2p 1d] ANO basis sets.

$^b$ Estimate for basis set incompleteness—see Table II.

$^c$ Reference 18.

$^d$ Reference 25.

$^e$ Reference 19. The error bars reflect the uncertainty in the heat of formation of CH$_3$.

$^f$ Reference 13.

$^g$ Reference 21.

$^h$ Reference 19. The error bars reflect the uncertainty in the heat of formation of NH$_2$. 